

Method	Fully Annotated Data	Use Play Data	No. Instructions in a Row (1000 chains)					
			1	2	3	4	5	Avg. Len.
Hulc	100%	✓	41.8%	16.5%	5.7%	1.9%	1.1%	0.67±0.10
MDT	100%	✓	61.7%	40.6%	23.8%	14.7%	8.7%	1.54±0.04
Spil	100%	✓	74.2%	46.3%	27.6%	14.7%	8.0%	1.71
Roboflamingo	100%	×	82.4%	61.9%	46.6%	33.1%	23.5%	2.47
Susie	100%	✓	87.0%	69.0%	49.0%	38.0%	26.0%	2.69
GR-1	100%	×	85.4%	71.2%	59.6%	49.7%	40.1%	3.06
3D Diff Actor	100%	×	93.8%	80.3%	66.2%	53.3%	41.2%	3.35±0.04
GR-MG w/o image	100%	×	91.0%	79.1%	67.8%	56.9%	47.7%	3.42±0.28
GR-MG w/o lang	100%	×	91.8%	79.8%	68.9%	58.1%	48.1%	3.46±0.04
GR-MG w/o prog	100%	×	94.1%	85.0%	75.2%	65.4%	56.3%	3.76±0.11
GR-MG (Ours)	100%	×	96.8%	89.3%	81.5%	72.7%	64.4%	4.04±0.03
GR-1	10%	×	67.2%	37.1%	19.8%	10.8%	6.9%	1.41±0.06
GR-MG w/o play data	10%	×	82.4%	60.8%	42.2%	28.7%	19.7%	2.33±0.04
GR-MG (Ours)	10%	✓	90.3%	74.5%	61.2%	47.4%	37.5%	3.11±0.08